# The economic geography of Europe and the role of regional policy

Julian Hinz

# The Economic Geography of Europe and the Role of Regional Policy*

Julian Hinz†

June 2012

Supervisor: Matthieu Crozet
Second Reader: Lionel Fontagné

**Abstract**

In this thesis I set up a standard New Economic Geography model and estimate it in the regional context of the European Union. The analysis underlines the clear core-periphery structure of Europe, but also identifies forces that hint at a catching-up of lesser developed peripheral regions. While regions that are close to the geographic center on average have a much higher market access, regions far from the center can improve their market access over the time period of 1999 - 2009 relative to their initial position. I estimate and evaluate the impact of European Union Cohesion Policy on this process and do not find the positive developments to be caused by or connected to the financial facilities of the European Union Regional Policy.

*I thank Matthieu Crozet for supervising this Master thesis and for his help. I am grateful to Sandy Dall'erba and Julie le Gallo for their help sharing and commenting on data on European Regional Policy. I thank Holger Breinlich for further comments on his research. I am thankful for the invaluable help of Rose von Richthofen and Nilofar Saidi.

†Paris School of Economics & Université Paris 1 Panthéon-Sorbonne

# 1 Introduction

> "The Community shall have as its task [...] to promote throughout the Community a harmonious, balanced and sustainable development of economic activities [...], the raising of the standard of living and quality of life, and economic and social cohesion and solidarity among Member States."
> — Article 2 of the Treaty establishing the European Community (1957)

The European Union has had a profound impact on the Economic Geography of the European Continent. The creation of the community in 1958 with the Treaty of Rome (1957) marked the starting point of an ever closer political and economic integration of a growing number of countries. In 2012, with now 27 member states, the political entity has more than half a billion inhabitants speaking in 23 official languages, that account for approximately 20% of the world's GDP (European Commission, 2008a).

Although the community has evolved into a highly integrated union, economic and social disparities have remained high. 43% of the European Union's economic output is generated in just 14% of its territory (European Commission, 2008c), particularly in the core situated in the geographic center of Europe. The peripheral regions are characterized by on average higher unemployment rates and a lower GDP per capita, emphasizing the strong core-periphery structure.

The economic integration of Europe has been at the heart of the entire integration process. The removal of borders and the introduction of one market for goods, the possibility for citizens of the EU to roam, live and work freely in any member country are known to have a vast impact on the economic geography of the continent (Crozet & Lafourcade, 2009, p. 79). The introduction of a single currency in wide parts of the political union, further reducing barriers for the internal market, has surely contributed to this end. In the context of the monetary integration and in light of Mundell (1961)'s criteria for an optimal currency area, Dall'erba (2008) recognizes that when "countries loos[e] control over monetary policy, cohesion policies [are] sometimes seen as the only tool to cushion asymmetric shocks and structural problems."

Since the Treaty of Rome in 1957, the European Community and later the European Union has the task "to promote throughout the Community a harmonious development of economic activities, a continuous and balanced expansion, an increase in stability, an accelerated raising of the standard of living and closer relations between the States belonging to it" (Treaty of Rome, 1957, Article 2). With establishing the *Structural Funds* and later the *Cohesion Fund*, this task received financial firepower.

EU regional policy has the aim to promote cohesion in economic, social and territorial terms for regions in the European Union. The size of the allocated funds is regularly about a third of the entire EU budget (European Commission, 2008b). The funds are allotted via different mechanisms to account for different objectives. In the period of 2000 - 2006 the two most important among others were defined as (European Communities, 2004):

*Objective 1*, for which roughly 80 percent of funds were transferred to regions, was the aim in NUTS2 regions where the GDP per capita is less than 75 percent of the EU25 average. Almost all regions of the new member states were covered by this objective. The primary obstacles to prosperity in these regions were assumed to be a general low level of investment, high unemployment and the lack of basic infrastructure.

*Objective 2*, for which about 16 percent of the available funds were allotted to, was the aim in the regions that were facing a structural change in previous key sectors, most notably declines in industrial activity and other traditional activities, leading to unemployment.

The *Cohesion Fund* provided another source of infrastructure funding and co-financing for countries that had a GDP of less than 90% of the EU average, covering similar areas as *Objective 1.*

In prior programming periods from 1989 - 1993, 1994 -1999 the aims were named differently and varied in numbers, the general objectives were of course closely related to the current ones. In the current period of 2007 - 2013, Objective 1 has been renamed the *convergence objective* and Objective 2 is part of the *competitiveness and employment objective* (European Commission, 2008a). The allocated funds are aimed at promoting economic growth and improved competitiveness on a regional level. This is to be achieved via several channels, most notably in infrastructure, human capital formation and social inclusion.

These policies have naturally been subject to economic research. Most of the econometric evaluations of European Union Regional Policy and its economic impact are situated in the Economic Growth literature, employing very different techniques leading to no coherent picture of the results (Dall'erba & Le Gallo, 2008). Among the notable contributions are Rodriguez-Pose & Fratesi (2004), employing a cross-sectional panel data analysis finding spending on infrastructure through Objective 1 funds to have an insignificant effect and only investments in human capital and education to have a medium run significant positive return. Becker *et al.* (2010) exploit the aforementioned threshold of a GDP below 75% of the average GDP for the eligibility for Objective 1 regions to identify causal effects of the receipt of funds on economic growth in the respective regions. Other authors, such as Dall'erba & Le Gallo (2008) and Gallo *et al.* (2011) employ spatial econometric techniques to account for spatial linkages in the estimations, however also reporting at best mixed results.

The main strand in economic research that is concerned with spatial features is of course Economic Geography, analyzing the roots and effects of agglomeration economies. The core-periphery concept, of apparent interest when studying the case of Europe, has received wide attention at least since the seminal contribution of Krugman (1991), a turning point for the henceforth *New* Economic Geography, emphasizing the interplay between proximity to markets for demand and supply, prices for input factors and trade costs.

The notion of market access has created a burgeoning literature with a wide range of topics. Hanson (1998, 2005) and Redding & Venables (2004) were successful in porting the theoretical advances to empirical tests. Redding & Venables (2004)'s econometric approach has been particularly influential, deriving a structural equation that relates wages to the proximity to demand and supplier markets, and achieving to explain a large part in the international variation of income with geographic features. Head & Mayer (2011) extend this analysis over a long period of time to find a long-term impact of market access on the economic development of countries.

Hanson (1998, 2005)'s work focusses on the subnational level, estimating the effects of market access on wage differentials for US counties. Other research applying this theoretical framework to country-level data include Brakman *et al.* (2004), who study the case of Germany.

A different strand of the economic geography literature exploits historical coincidences as a form of natural experiments: Redding & Sturm (2008) use the German partition and reunification between the end of World War 2 and 1990 showing the importance of market access through a decline of population in border areas, as suggested by the theory. Davis & Weinstein (2002) test the durability of economic geography against external shocks, analyzing the population distributions before and after allied bombings of World War 2 on Japanese cities, finding a strong persistence of established economic geographical patterns.

In the European context, a number of works stand out. Head & Mayer (2004) use a similar framework as Redding & Venables (2004) to develop a model of location choice for firms, and show using data on Japanese investments in Europe that investments are made "where the markets are", focussing the research on the backward linkages to demand sources. Crozet (2004) sheds light on the forward linkages, developing and estimating a model of individual location choice with interregional migration data for five European countries. Combes & Overman (2004) provide a survey over literature and open questions over for the specific case of Europe.

Breinlich (2006) follows closely the methodology and focus of research of Redding & Venables (2004) and applies the framework to European regions, identifying human capital accumulation as a channel through which *market access* influences wages. Head & Mayer (2006), also focussing on European regions, modify the approach slightly and are looking at wage and employment *responses* to proximity to markets for different industries.

This paper also makes use of the Redding & Venables (2004) methodology and through its European focus is closely related to Breinlich (2006) and Head & Mayer (2006). It extends the analysis to all 248 regions of the 25 member states of 2006 that were located in continental Europe. Analyzing the determinants for the economic geography of Europe, I estimate the impact of Objective 1 (and Cohesion Fund) and Objective 2 facilities on the dynamics and determinants of the distribution of income.

The paper is organized as follows: In section 2 I set up the theoretical framework in which I am to analyze the Economic Geography of Europe. The following sections are concerned with the empirical testing of this model, where the *trade equation* is estimated and market access is calculated (Section 3) and these results then used in the estimation of the *wage equation* (Section 4). In section 5 I draw the link to the policy dimension and estimate the role of European Union Regional Policy in the determinants for change of the distribution of income across Europe. Section 6 finally concludes.

## 2 Theoretical Framework

In this section I set up a standard New Economic Geography model, modified for the purpose of an analysis of regional wage differences. The framework is fundamentally based on Fujita *et al.* (1999), and comparable to Redding & Venables (2004), while highlighting the regional approach similar to Head & Mayer (2006) and Breinlich (2006). I loosely follow the notation of Combes *et al.* (2008) for its ease of use and readability.

The model is made up by $R$ regions, where in each region $r$ there are $n_r$ firms that produce their variety of a good $q_r$. The consumer in region $r$, who derives his income $Y_r$ from wages $w_r$ and rent on capital $x_r$, and spends a fraction $\mu_r$ on the produced products from region $r$ and all other regions $s \in R$. In the following, $r$ is to be perceived as the domestic region, whereas $s$ is a foreign region, simplifying the understanding of the model.

## 2.1 Supply Side

As usual, the model consists of a supply and demand side. On the supply side, for the firm in region $r$ to produce one unit of its product, it is faced with a function

$$q_r = AL_r^{\alpha} X_r^{\beta} - a \qquad (1)$$
$$\text{with } \alpha + \beta = 1$$

where in the Cobb-Douglas function-like component $L_r$ is the labor input, $\alpha$ the associated labor share in income, $X_r$ the input of capital and $\beta$ the share of capital in income. $A$ is a standard technological parameter and $a$ is a fixed input which determines the degree of increasing returns to scale, both are assumed to be identical across all regions.

The firm maximizes the production of its variety $q_r$ while operating under a cost constraint. The optimization problem is therefore:

$$\max_{L_r, X_r q_r} = AL_r^{\alpha} X_r^{\beta} - a$$
$$\text{s.t. } C_r = w_r L_r + x_r X_r$$

The optimization leads to the cost function

$$C_r = w_r^{\alpha} x_r^{\beta} (q_r + a)$$

For the firm this implies a profit

$$\Pi_r = \sum_{s=1}^{R} p_r q_{rs} - w_r^{\alpha} x_r^{\beta} (q_r + a)$$

This introduces the notion of distance, as $q_{rs}$ is the quantity of a variety $q_r$ that is produced in region $r$ and then sold in (or exported to) region $s$. Splitting the cost into fixed and marginal cost this becomes

$$\Pi_r = \sum_{s=1}^{R} p_r q_{rs} - m_r q_r - F_r \qquad (2)$$
$$\text{with the marginal cost } m_r = w_r^{\alpha} x_r^{\beta} \qquad (3)$$
$$\text{and the fixed cost } F_r = a w_r^{\alpha} x_r^{\beta} \qquad (4)$$

The operating profit of a firm in region $r$ exporting to region $s$ in equilibrium is then

$$\pi_{rs}^{*} = (p_r^{*} - m_r) \tau_{rs} q_{rs}^{*} \qquad (5)$$

with an equilibrium mill price $p_r^*$, the equilibrium quantity $q_{rs}^*$ and trade cost $\tau_{rs} \geq 1$. The trade costs are *ad valorem*, or so-called *iceberg* trade costs, meaning a fraction of $\frac{\tau_{rs}-1}{\tau_{rs}}$ of the shipped goods "melts" on the way from region $r$ to $s$. Under increasing returns to scale and imperfect competition, the firm establishes an equilibrium price for the consumer[1] of

$$p_{rs}^* = \tau_{rs} p_r^* = \tau_{rs} m_r \left( \frac{\sigma}{\sigma - 1} \right) \tag{6}$$

receiving a markup of $\frac{\sigma}{\sigma-1}$ over the marginal cost. Substituting (6) into (5) yields

$$\pi_{rs}^* = m_r \frac{\tau_{rs} q_{rs}^*}{\sigma - 1} \tag{7}$$

### 2.2 Demand side

On the consumer side, the equilibrium demand $q_{rs}^*$ for goods in region $s$ from region $r$ is derived from the consumer's utility maximization. The agent is assumed to have a preference for consuming all available varieties from all regions $R$, characterized by a CES utility function:

$$\max_{q_{rs}} U_s = \left( \sum_{r=1}^{R} n_r q_{rs}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{8}$$
$$\text{s.t.} \sum_{r=1}^{R} n_r p_{rs} q_{rs} = \mu_s Y_s$$

As described above, $Y_s$ is the income in region $s$, $\mu_s$ the share of income devoted to the consumption of the considered good, $\sigma$ the elasticity of substitution, $n_r$ the number of varieties in region $r$. The optimization leads to

$$q_{rs}^* = p_{rs}^{-\sigma} \frac{\mu_s Y_s}{\sum_{r=1}^{R} n_r p_{rs}^{1-\sigma}} \tag{9}$$

which can be further simplified by defining a price index $P_s$

$$q_{rs}^* = p_{rs}^{*-\sigma} \mu_s Y_s P_s^{\sigma-1} \tag{10}$$

$$\text{with } P_s = \left( \sum_{r=1}^{R} n_r p_{rs}^{*1-\sigma} \right)^{\frac{1}{1-\sigma}} \tag{11}$$

The equilibrium quantity exported from region $r$ to $s$ is therefore negatively dependent on its price, and positively dependent on the share of income devoted to consumption of this good, discounted by the price level in the importing region.

[1] Essentially the *CIF* price.

Multiplying (10) on both sides with $n_r p_{rs}$ yields

$$n_r p_{rs}^* q_{rs}^* = n_r p_{rs}^{*1-\sigma} \mu_s Y_s P_s^{\sigma-1} \tag{12}$$

which I will call the *trade equation* in accordance with previously mentioned authors using a similar setup and following Redding & Venables (2004). This equation will be estimated in section 3.1.

Returning to the firms' profit function (2), summing over all exports, that is to say all operating profits (7), from region $r$ and subtracting the fixed costs (4) then yields a total profit of

$$\begin{aligned} \Pi_r^* &= \sum_s^R \pi_{rs}^* - F_r \\ &= \sum_s^R m_r^{1-\sigma} \tau_{rs}^{1-\sigma} \sigma^{-\sigma} (\sigma - 1)^{\sigma-1} \mu_s Y_s P_s^{\sigma-1} - F_r \end{aligned} \tag{13}$$

Rearranging terms and repackaging the constants into $c = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}$ yields

$$\Pi_r^* = c m_r^{1-\sigma} MA_r - F_r \tag{14}$$

with calling the sum over income of importing regions $s$ discounted by respective price indexes and trade costs the *market access* following the notation of Redding & Venables (2004):[2]

$$MA_r = \sum_{s=1}^R \mu_s Y_s P_s^{\sigma-1} \tau_{rs}^{1-\sigma} \tag{15}$$

Assuming zero profit in the long-run, the profit equation (14) can be rearranged to

$$\begin{aligned} \Pi_r^* &= c m_r^{1-\sigma} MA_r - F_r = 0 \\ \Leftrightarrow m_r &= \left(\frac{cMA_r}{F_r}\right)^{\frac{1}{\sigma-1}} \end{aligned}$$

Inserting the marginal cost $m_r$ and fixed cost $F_r$ from the optimization program of the

[2]Head & Mayer (2011) provide a discussion over naming this term *market access* or *real market potential*

firm, equations (3) and (4), and solving for $w_r$ yields

$$\begin{aligned}\left(\frac{cMA_r}{F_r}\right)^{\frac{1}{\sigma-1}} &= w_r^{\alpha} x_r^{\beta} \\ w_r &= \left(\frac{c}{a}\right)^{\frac{1}{\alpha\sigma}} MA_r^{\frac{1}{\alpha\sigma}} x_r^{-\frac{\beta}{\alpha}} \end{aligned} \tag{16}$$

which is the familiar *wage equation* named by Fujita *et al.* (1999).

So far the approach is purely static. As described above, one of the aims of this paper is to analyze the *change* over time, in particular the *change of wages*. I therefore introduce a time dimension and divide observations in time $t$ by the previous one $t-1$:

$$\frac{w_{r,t}}{w_{r,t-1}} = \left(\frac{c_t a_{t-1}}{c_{t-1} a_t}\right)^{\frac{1}{\alpha\sigma}} \left(\frac{x_{r,t}}{x_{r,t-1}}\right)^{-\frac{\beta}{\alpha}} \left(\frac{MA_{r,t}}{MA_{r,t-1}}\right)^{\frac{1}{\alpha\sigma}}$$

The exponents $\alpha$, $\beta$, the shares of factor returns, are assumed not to change over time.Taking the logarithm, I obtain the first differences of the wage equation, describing the growth rate of the remuneration for the immobile factor:

$$\begin{aligned}\ln w_{r,t} - \ln w_{r,t-1} = \frac{1}{\alpha\sigma}(\ln c_t - \ln a_t - \ln c_{t-1} + \ln a_{t-1}) - \frac{\beta}{\alpha}(\ln x_{r,t} - \ln x_{r,t-1}) \\ + \frac{1}{\alpha\sigma}(\ln MA_{r,t} - \ln MA_{r,t-1})\end{aligned}$$

which I abbreviate with:

$$\Delta \ln w_{r,t} = \frac{1}{\alpha\sigma}(\Delta \ln c_t - \Delta \ln a_t) - \frac{\beta}{\alpha}\Delta \ln x_{r,t} + \frac{1}{\alpha\sigma}\Delta \ln MA_{r,t} \tag{17}$$

The change in the remuneration for the immobile factor, namely wages, is therefore positively dependent on the change in the elasticity of substitution, negatively dependent on the change of the degree of IRS and a change in the remuneration of mobile capital, while a change in market access increases wages. The former three appear reasonable, particularly any relative change in the remuneration for the mobile input, capital, means relatively less remuneration for the immobile input ceteris paribus and a higher degree of increasing returns to scale, essentially higher fixed costs, implies a lower number of firms resulting in lower wages. The latter is intuitive as well: greater market access implies a more profitable environment for the firm, thus the ability to pay higher wages.

# 3 Estimation of the Trade Equation

In the previous section 2, two equations have been highlighted: the *trade equation* (12) and the *wage equation* (16). These two are fairly standard in the literature, while less common is the first difference of the wage equation, equation (17), describing the *growth rate of wages.*

In order to estimate the model, two stages are necessary. In the first stage, I estimate the trade equation, essentially a standard gravity equation, with two different but related approaches, which I call *RV method* after Redding & Venables (2004) and *HM method* after Head & Mayer (2004) as coined by Paillacar (2009). Both methodologies make use of fixed effects for importers and exporters to obtain estimates for the construction of the market access $MA_r$ in section 3.5, but differ in the data and trade costs variables used, which will be explained below in section 3.2.

In a second stage in sections 4 and 5, I use these constructed variables for the estimations of equations (16) and (17), the *wage equation* and the respective *growth rate of wages.*

## 3.1 Econometric specification

The trade equation is estimated in a similar fashion as Redding & Venables (2004), using fixed effects for importer and exporter to capture what Anderson & van Wincoop (2003) call *multilateral resistance term.*

Recalling equation (12) from section 2 and separating the *CIF* price into *FOB* price and trade costs, the trade equation becomes

$$n_r p_{rs}^* q_{rs}^* = \left(n_r p_r^{*1-\sigma}\right)\left(\tau_{rs}^{1-\sigma}\right)\left(\mu_s Y_s P_s^{\sigma-1}\right) \tag{18}$$

The left hand side is the total value of all varieties shipped from region $r$ to $s$, in other words the total value of exports. The first term on the right hand side is exporter specific, the number and local prices of domestic varieties. The second term is trade costs between the two regions. The third term finally is importer specific, the importer's share of income allotted to the considered good and the foreign price level. Condensing these attributes and introducing a time dimension yields

$$X_{rs,t} = (i_{r,t})(\tau_{rs,t}^{1-\sigma})(j_{s,t}) \tag{19}$$

where $X_{rs,t}$ is the bilateral trade flow from region $r$ to $s$ at time $t$ and $i_{r,t}$ and $j_{s,t}$ are the exporter and importer-specific features. The latter will be used to estimate the market capacity, the expenditure discounted by price level, of the locations. Taking logs yields

$$\ln X_{rs,t} = \ln i_{r,t} + (1-\sigma)\ln \tau_{rs,t} + \ln j_{s,t}$$

I estimate this equation in both the *RV method* and the *HM method* then with

$$\ln X_{rs,t} = \theta_t + \eta_{1,r,t} EX_r + \eta_{2,s,t} IM_s + (1-\sigma)\,\eta_{3,t} TC_{rs}$$

$EX_r$ is the dummy for the exporting region $r$, $IM_s$ is the dummy for the importing region $s$ and $TC_{rs}$ contains a number of variables responsible for transport cost. In this case, trade costs will be assumed to depend on distances $DIST_{rs}$ and a shared official language $LANG_{rs}$. In the *HM method*, I also include a dummy $BORDER_{rs}$ if a trade flow leaves the country, i.e. when exporter and importer regions are not located in the same country, as intra-country trade between regions is expected to reach different magnitudes than other trade flows.[3] On the other hand, in the *RV method* I follow Redding & Venables (2004) and include a variable $CONTIG_{rs}$ if two regions are contiguous. The estimated equation is then

$$\ln X_{rs,t} = \theta_t + \eta_{1,r,t} EX_r + \eta_{2,s,t} IM_s + \zeta_{3,t} DIST_{rs} + \zeta_{4,t} LANG_{rs} + \zeta_{5,t} BORDER_{rs} + \epsilon_{rs,t} \tag{20}$$

$$\ln X_{rs,t} = \theta_t + \eta_{1,r,t} EX_r + \eta_{2,s,t} IM_s + \zeta_{3,t} DIST_{rs} + \zeta_{4,t} LANG_{rs} + \zeta_{5,t} CONTIG_{rs} + \epsilon_{rs,t} \tag{21}$$

for the *HM method* and *RV method* respectively, where in both $\zeta_i = (1-\sigma)\,\eta_i$.

## 3.2 Two Approaches for Estimation of the Trade Equation

As described, I estimate the trade equation applying two different approaches. Theoretically, both approaches are fully consistent with the framework set up in section 2. Empirically however, as will be displayed in section 3.4, the two methodologies deliver slightly differing results.

Redding & Venables (2004) estimate the trade equation with bilateral trade data from all countries. Through the importer fixed effect, the estimation yields an estimate for the income discounted by price level for all locations, without the need for internal trade data. Paillacar (2009) coins this the *RV method*. While this technique certainly has the advantage of ridding the estimation of the need for internal trade data, it has a clear disadvantage too: it does not allow for a border effect, which then can potentially bias the estimates for other assumed trade costs.

Head & Mayer (2004) proceed differently: they include a dummy variable for international trade flows, creating the need for internal trade data. The dummy is designed

[3] See Anderson & van Wincoop (2004) for a survey on this topic.

to capture all costs associated with the crossing of a national border, "comprising [a] home bias in consumer preferences and government procurement, differential technical standards, exchange rate uncertainty, and imperfect information about potential trade partners" (Head & Mayer, 2004, p. 962), and therefore should have a negative sign. Paillacar (2009) names this approach the *HM method*. Breinlich (2006), also applying the original framework by Redding & Venables (2004) to a select number of European regions, adapts the approach further in exclusively using trade flows originating from the European Union, therefore requiring only European export and production data. In my estimation using the *HM method*, I proceed similar to Breinlich and use exclusively data on exports and production from the 25 member states of the European Union before the enlargement of 2007.

Except for the treatment of internal trade flows, all variables are the same. The results in section 3.4 emphasize the technical differences. As the exclusive use of European data for the *HM method* significantly reduces the data sample, I again follow Breinlich (2006) and pool data to four periods: 1999−2001, 2002−2004, 2005−2007, 2008−2009. As Breinlich notes, this also has the positive effect of reducing short-term fluctuations in the sample.

### 3.3 Data

For the estimation of the trade equation, equations (20) and (21), data on bilateral trade flows and trade costs is necessary. Unfortunately, although EUROSTAT and affiliated agencies such as ESPON are providing rich regional data in many aspects, trade data in a regional detail is unavailable on a European level. While for some European countries more fine-grained data is available and used in related research, see e.g. Combes *et al.* (2005) using regional trade flows between French regions, complete coverage for all EU member states is only available in national figures.[4]

Due to this constraint, I have to resort to national accounts on trade flows. The most complete database for trade flows I find to be UN Comtrade, and I use the cleaned and completed BACI version from CEPII (Gaulier & Zignago, 2010). The highly disaggregated product-level data is reaggregated to the sector level using a concordance table from HS6 to ISIC Rev. 2.[5] As previous authors I use manufacturing exports for the estimations, hence aggregate at the ISIC Rev. 2 top-level 3 code. BACI is accompanied by a smaller dataset that enables a distinction between true zeros and missing values. Like Redding & Venables (2004) I add 1 for all zero trade flows to prohibit $log(0)$s in the regression. As all figures are reported in thousands of dollars, a

[4] Related research that makes use of regional trade data from countries in other part of the world are among others Wolf (2000) and Anderson & van Wincoop (2003) using data from US states, and Hering & Poncet (2010), using regional Chinese trade data.

[5] *http://www.macalester.edu/research/economics/page/haveman/Trade.Resources/tradeconcordances.html*

negligible \$1.000 trade flow is therefore assumed to take place. The dataset treats the countries of Luxembourg and Belgium as one. Luxembourg is hence assumed to be a region situated in Belgium.

As previously elaborated, when estimating the trade equation with the *RV method*, no information on internal trade flows is necessary. However, for the *HM method*, this is the case. Like Head & Mayer (2004, 2006) and Breinlich (2006), I follow Wei (1996) and calculate internal trade as the difference between manufacturing production and exports. Data on manufacturing production is derived from the OECD STAN database.

As for the required information on trade costs, I make use of two databases provided by CEPII, the *Distances* database (Mayer & Zignago, 2011) and the *Gravity* database (Head *et al.* , 2010), using information on the location of capitals for the calculation of distances, shared official languages and common borders with trading partners. However, as this paper is concerned with European regions, I have created an analogous database for European NUTS2 region of the EU25 using the same strategy as detailed in Mayer & Zignago (2011) applied to European regions. I include 248 regions that lie on the European continental shelf.[6]

Following Paillacar (2009), for distances I use simple great circle distances. Head & Mayer (2002) note that this overestimates the home bias in trade and propose a population-weighted distance measure. Paillacar however finds that ad hoc great circle distances do not significantly alter the estimations, while recognizing that data on internal population is highly incomplete. This is even more true on a regional level, where the classification of regions is already somewhat ambiguous.[7] For internal distances a standard approach is to use the area of the region and assuming it to have a disk-like shape, for which the average distance is calculated as $DIST_{rr} = \frac{2}{3}\sqrt{area_r/\pi}$.

### 3.4 Results from Estimation of the Trade Equation

Estimating the trade equations (20) and (21), the results between the two approaches differ in the expected form. Table 1 shows the results of the estimation with the *RV method*. Distance, a common language, and a shared border (contiguity) together can explain about 76% of variation in the data. All coefficients have the expected sign and are economically and statistically highly significant for all years.[8] The coefficients on

[6] I exclude any overseas territories or departements, Portuguese and Spanish islands in the Atlantic Ocean, Spanish enclaves in Northern Africa, and Gibraltar. See Appendix A for further information.

[7] The *Nomenclature of Units for Territorial Statistics* for European regions classifies by population size: NUTS2, which is used in this paper, is a body of land that is inhabited by a population of size between 800.000 and 3.000.000 and is mainly based on institutional divisions in member states.

[8] As annual data has been grouped into four periods of two to three years, standard errors are clustered around importer-exporter pairs. Significance is reported with t-statistics based on cluster-robust standard errors.

Table 1: Trade equation RV Method 1999 - 2009

| Regressor | 1999 - 2001 (1) | 2002 - 2004 (2) | 2005 - 2007 (3) | 2008 - 2009 (4) |
|---|---|---|---|---|
| Distance | -1.63011*** | -1.639549*** | -1.648194*** | -1.59885*** |
| | (-138.1693) | (-140.5560) | (-136.8687) | (-114.167) |
| Common Language | 1.06157*** | 1.070790*** | 1.106022*** | 1.02921*** |
| | (42.8597) | (43.9066) | (45.6562) | (34.942) |
| Contiguity | 1.28030*** | 1.204572*** | 1.137786*** | 0.90830*** |
| | (17.5574) | (16.5814) | (14.9626) | (12.054) |
| Adjusted $R^2$ | 0.7602 | 0.7675 | 0.771 | 0.7675 |
| Observations | 97520 | 98185 | 97996 | 64887 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$, cluster-robust t-statistic in round brackets (standard errors clustered around country pairs). Dependent variable is the logarithm of trade flow, the independent variables are logarithm of distance and binary indicators for a common language and contiguity of countries.

distance are slightly higher for all periods than in Redding & Venables (2004)[9] and marginally *increase* until the fourth period, which is also inline with the literature.[10] The importance of contiguity between trading partners is higher than in Redding & Schott (2003) and Redding & Venables (2004) but steadily decreasing. A common language between trading partners, not included in the original Redding & Venables (2004) estimation, displays no particular trend and range marginally lower than the reported values from Breinlich (2006).

Tables 2 and 3 show the results for the estimation of the trade equation with the *HM method*. While all coefficients in both tables have the expected sign, the magnitude differs tremendously. For all 25 countries of the European Union the coefficient on distance is much larger than for the old member states before the 2004 enlargement, while the border coefficient is much lower for the estimation with all member states compared to the selection of the EU15. This result is somewhat puzzling, as it implies much lower trade reductions when crossing a border in the new member state countries compared with the old member states. The issue is most likely rooted in the construction of internal trade data. When exclusively looking at the estimation for the countries that were members of the EU before the enlargement of 2004, the coefficients are to a large extent in line with previous research: the implied trade reductions by crossing a national border mirror Breinlich (2006)'s findings, with values decreasing from $-72\%$ in the first period to $-64\%$ in the third period. Note however that

[9] Compare (Redding & Venables, 2004, p. 62, Table 1, Column 1) for the same methodology used.
[10] See Disdier & Head (2008) for a survey on the topic of the high negative impact of distance on trade.

Table 2: Trade equation HM Method 1999 - 2009

| Regressor | 1999 - 2001 (1) | 2002 - 2004 (2) | 2005 - 2007 (3) | 2008 - 2009(4) |
|---|---|---|---|---|
| Distance | -1.77024*** | -1.72813*** | -1.72492*** | -1.71556*** |
| | (-37.7113) | (-39.0121) | (-34.2575) | (-33.0368) |
| Common Language | 1.09646*** | 1.03010*** | 1.09456*** | 1.20567*** |
| | (15.5570) | (14.3467) | (16.2610) | (14.7680) |
| Border | -0.43596* | -0.36505* | -0.19536 | -0.18334 |
| | (-1.9220) | ( -1.8177) | (-0.8426) | (-0.7047) |
| impl. trade reduction | 0.35 | 0.31 | 0.18 | 0.17 |
| Adjusted $R^2$ | 0.8637 | 0.8696 | 0.8674 | 0.88 |
| Observations | 14777 | 14776 | 14775 | 9847 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$, cluster-robust t-statistic in round brackets (standard errors clustered around country pairs). Dependent variable is the logarithm of trade flow, the independent variables are logarithm of distance and binary indicators for a common language and the crossing of a border. Implied trade reduction is $1 - \exp(border)$.

Breinlich (2006) and Head & Mayer (2006) allow border effects to vary across different countries.[11] The coefficients are only statistically significant for the first two periods. In both estimations of EU25 and EU15 a common language plays an economically and statistically significant role in the determination of the size of trade flows, although the coefficient varies across time. The estimation of the trade equation with the *HM method* explains between $86\% - 90\%$ of the variation of trade flows. Although coefficients vary substantially between the estimations for the EU25 and EU15, to ensure consistency those coefficients of the estimation with all 25 countries of the European Union are used for the calculation of the market access.

In the analysis below, mostly the results from the *HM method* will be used, as the information on internal trade flows appears to be extraordinarily important in the context of regions, as will be shown below. However, the results from the *RV method* are overall similar and provide a good contrast to be used to underpin the general findings and highlight the differences rooted in the construction of the two methods.

### 3.5 Computation of Market Access

Having obtained estimates for trade costs and importer fixed effects, I can proceed to compute the market access for all regions. However, the trade equation, as described above, is estimated with *national* data. To make the step to *regional* market access, one

[11] Letting border effects vary across countries in my case led to enormous variations in the later calculated market access. I therefore resorted to assuming a common effect, although Chen (2004) finds significant differences for the effect for European countries.

Table 3: Trade equation HM Method 1999 - 2009 for EU15 only

| Regressor | 1999 - 2001 (1) | 2002 - 2004 (2) | 2005 - 2007 (3) | 2008 - 2009 (4) |
|---|---|---|---|---|
| Distance | -1.1965275*** | -1.203978*** | -1.175567*** | -1.181475*** |
| | (-22.4601) | (-23.6480) | (-19.0561) | (-18.6857) |
| Common Language | 1.1780656*** | 1.116936*** | 1.213514*** | 1.284977*** |
| | (17.6732) | (1.060122) | (19.4007) | (16.4424) |
| Border | -1.2741031* | -1.060122* | -1.037907 | -1.141489 |
| | (-6.1237) | ( -4.5934) | (-4.0140) | (-3.8505) |
| impl. trade reduction | -72 % | -65 % | -64 % | -68 % |
| Adjusted $R^2$ | 0.8954 | 0.9003 | 0.892 | 0.9059 |
| Observations | 8844 | 8844 | 8844 | 5898 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$, cluster-robust t-statistic in round brackets (standard errors clustered around country pairs). Dependent variable is the logarithm of trade flow, the independent variables are logarithm of distance and binary indicators for a common language and the crossing of a border. Implied trade reduction is $1 - \exp(border)$. Note that this estimation only includes data on EU15 countries.

critical assumption has to be made: all regions in a country share the same price-level. While this assumption is quite strong and disregards apparent differences (Combes *et al.* , 2008), it allows to split the estimated importer country fixed effect by regional expenditure shares:

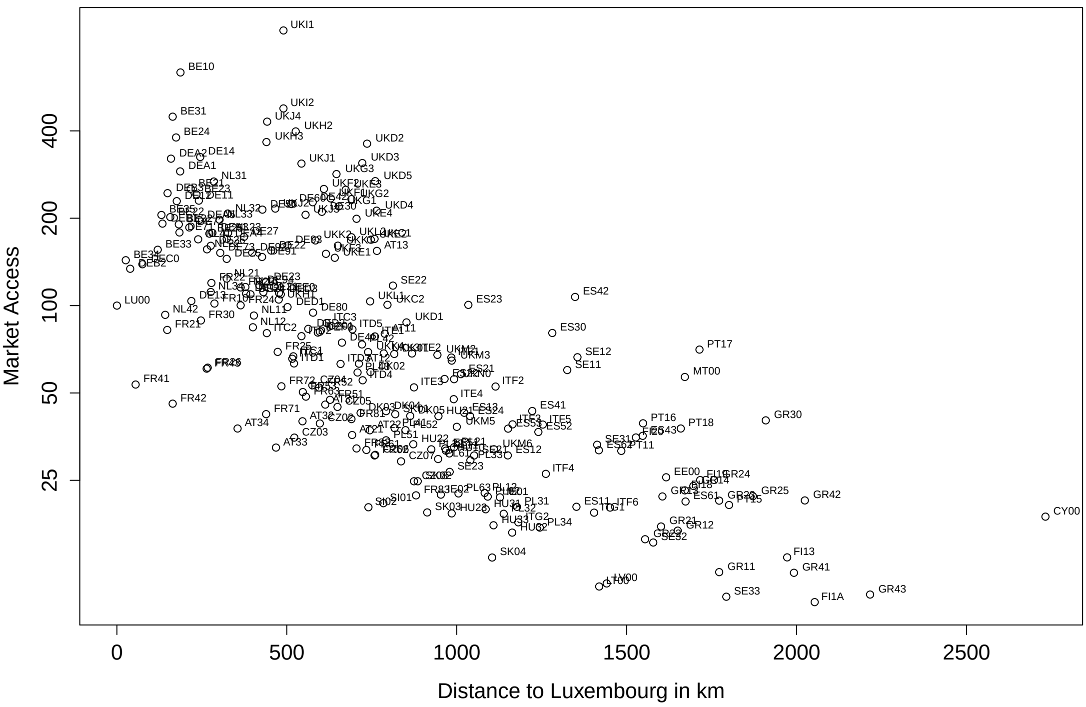
$$\exp(IM_s)^{\eta_{2,s}} = \frac{\mu_s Y_s}{\mu_S Y_S} \exp(IM_S)^{\eta_{2,S}} = \frac{\mu_s Y_s}{P_S^{1-\sigma}} \tag{22}$$

where region $s$ is situated in the importing country $S$, implying $P_s = P_S, \forall s \in S$. The trade equation is also estimated with national data on trade costs, so for the calculation of region market access, the regional data is used, relying on official regional languages and distances from regions to countries, other regions within the same country and internal distances as described in section 3.3.

Recalling equation (15), the market access will then be calculated using equation (19) accordingly for the respective time $t$ as:

$$\begin{aligned} MA_{r,t} &= \sum_{s=1}^{R} \mu_{s,t} Y_{s,t} P_{s,t}^{\sigma-1} \tau_{rs,t}^{1-\sigma} = \sum_{s=1}^{R} (j_{s,t})(\tau_{rs,t}^{1-\sigma}) \\ \Rightarrow MA_{r,t} &= \sum_{s=1}^{R} \Big( \exp(IM_s)^{\eta_{2,s,t}} (DIST_{rs})^{\zeta_{3,t}} \exp(LANG_{rs})^{\zeta_{4,t}} \exp(BORDER_{rs})^{\zeta_{5,t}} \Big) \end{aligned} \tag{23}$$

Figure 1: Distance to Luxembourg and Market Access in Period 1999 - 2001




*Note:* Distance to Luxembourg in kilometers and market access relative to Luxembourg (LU00 = 100). Market access calculated with HM method for period from 1999 - 2001.

for the *HM method* and

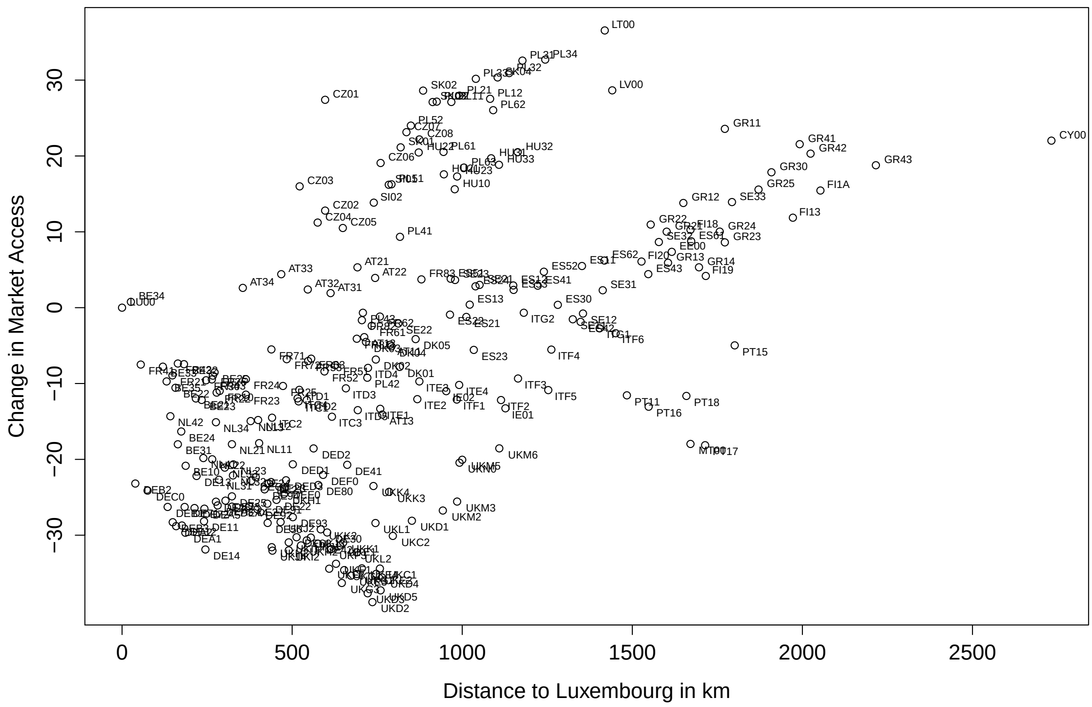
$$\Rightarrow MA_{r,t} = \sum_{s=1}^{R} \Big( \exp(IM_s)^{\eta_{2,s,t}} (DIST_{rs})^{\zeta_{3,t}} \exp(LANG_{rs})^{\zeta_{4,t}} \exp(CONTIG_{rs})^{\zeta_{5,t}} \Big) \tag{24}$$

for the *RV method*.

Figure 1 displays the resulting market access values of the 248 regions, using coefficients calculated with the *HM method*, for the period of 1999 - 2001 relative to their distance from Luxembourg. The trend is easily visible: the further a region is located from the geographic center of the European Union the lower is its market access. Particularly well positioned and blessed with a high market access seem to be the metropolitan regions of London (NUTS 2 codes UKI1 and UKI2) and Brussels (NUTS 2 codes BE10, BE31 and BE24).[12] Of particular interest with respect to the latter part of this paper, section 5, is figure 2. While less clear than the relationship between distance to the geographic center and market access, it appears that regions located far from the center

[12]See figure 5 in appendix C for the analogous plot for market access calculated with the *RV method*.

Figure 2: Distance to Luxembourg and Relative Change in Market Access 1999 - 2009




*Note:* Distance to Luxembourg in kilometers and change in market access from period 1999 - 2001 to period 2008 - 2009, relative to Luxembourg. Market access calculated with HM method.

improved their market access in the period of 2008 - 2009 relative to their initial market access in the period of 1999 - 2001 more than those regions located closer to the center.

The differences between the two techniques used to estimate the *trade equation* become apparent when looking at table 4. Market access calculated based upon coefficients from the *RV method* is muss less reliant on the domestic component of trade, while the corresponding figures for the *HM method* mirror the home bias in trade, with respect to both domestic region and country, captured by the border dummy in the estimation. Note also that the latter shows a clear trend for European regions in the composition of their market access: Europe and the rest of the world gain shares at the expense of the domestic region and country. In the *RV method* no such clear trend is visible, and while the order of importance is equivalent to the *HM method*, it seems to deliver less precise results.

Table 4: Composition of Market Access for all period using HM method and RV method

| *HM method* | '99 - '01 | '02 - '04 | '05 - '07 | '08 - '09 |
|---|---|---|---|---|
| Region | 12.36 | 11.88 | 11.53 | 11.63 |
| Country | 70.31 | 67.76 | 65.36 | 65.03 |
| Europe | 14.21 | 16.52 | 18.62 | 18.53 |
| World | 3.12 | 3.84 | 4.49 | 4.81 |
| *RV method* | '99 - '01 | '02 - '04 | '05 - '07 | '08 - '09 |
| Region | 7.92 | 8.19 | 8.21 | 7.93 |
| Country | 51.56 | 53.53 | 52.62 | 52.11 |
| Europe | 38.30 | 36.25 | 36.93 | 37.24 |
| World | 2.23 | 2.03 | 2.24 | 2.72 |

*Note:* Market access in top half calculated with the HM method, in bottom half calculated with RV method. Cells show average share of regional, national, European and worldwide component in market access for all 248 regions per time period.

# 4 Estimation of the Wage Equation

Having obtained estimates for market access, I can proceed to apply these values in the estimation of the wage equation (16) and the related change in wages (17). Recalling from section 2 the wage equation (16) reads:

$$w_r = \left(\frac{c}{a}\right)^{\frac{1}{\alpha\sigma}} x_r^{-\frac{\beta}{\alpha}} MA_r^{\frac{1}{\alpha\sigma}}$$

Taking logs and applying a time dimension yields

$$\ln w_{r,t} = \frac{1}{\alpha\sigma}\ln\left(\frac{c_t}{a_t}\right) - \frac{\beta}{\alpha}\ln x_{r,t} + \frac{1}{\alpha\sigma}MA_{r,t}$$

Recall that $a_t$ can vary over time but is invariant across regions. This is then estimated as

$$\ln w_{r,t} = \theta_t + \frac{1}{\alpha\sigma}\ln MA_{r,t} + \epsilon_{r,t} \tag{25}$$

The renumeration for the mobile input, $x_{r,t}$, is captured in the error term, while the constant $\theta_t$ captures the time-variant $a_t$ and $c_t$. All cross-regional differences are thus assumed to be captured in the residual, which is a quite strong hypothesis (Redding & Venables, 2004). In section 4.2 I therefore conduct several robustness tests.

Wages, the remuneration for the immobile factor, is proxied with GDP per capita, following numerous related papers. As I am calculating the market access over periods, I also take the mean GDP per capita for the respective time periods. Regional GDP data comes from EUROSTAT.

## 4.1 Baseline Estimation

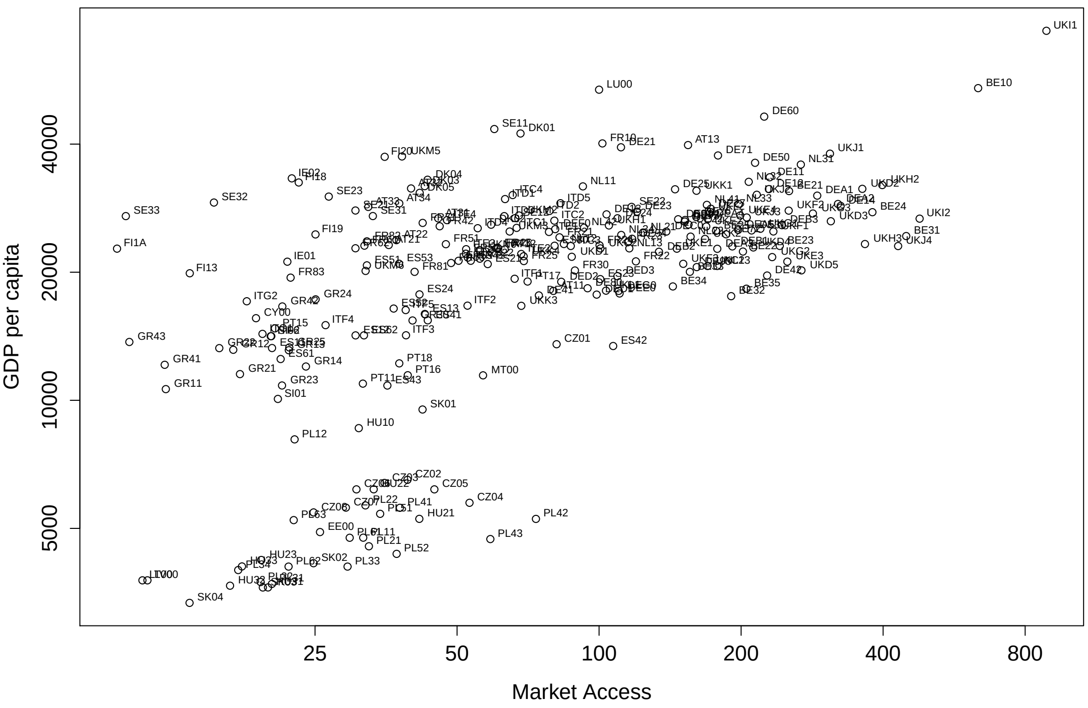
Table 5: Baseline estimation wage equation 1999 - 2009

| Regressor | ln(GDPpc) (1) | ln(GDPpc) (2) | ln(GDPpc) (3) | ln(GDPpc) (4) |
|---|---|---|---|---|
| ln(MA RV) | 0.25744*** | | | |
| | (17.1927) | | | |
| ln(MA HM) | | 0.345281*** | | |
| | | (18.9450) | | |
| ln(MA) P1 | | | 0.30239*** | 0.38873*** |
| | | | (9.4554) | (10.7178) |
| ln(MA) P2 | | | 0.267724*** | 0.35773*** |
| | | | (8.6643) | (9.6623) |
| ln(MA) P3 | | | 0.237804*** | 0.32963*** |
| | | | (8.4723) | (9.3684) |
| ln(MA) P4 | | | 0.212776*** | 0.28435*** |
| | | | (7.9523) | (8.2021) |
| Period Dummies | yes | yes | yes | yes |
| Adjusted $R^2$ | 0.2702 | 0.2965 | 0.2722 | 0.2978 |
| Observations | 992 | 992 | 992 | 992 |
| Year | 1999 - 2009 | 1999 - 2009 | 1999 - 2009 | 1999 - 2009 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$, t-statistic in round brackets. Due to the nature of the regressor being itself generated from a prior regression, t-statistics should be reported based on bootstrapped standard errors, which however could not be done. The independent variable is the logarithm of the mean of GDP per capita in the respective period, as described in the text. The regressors in columns (1) and (2) are the logarithms of the market access calculated with the RV method in column (1) and HM method in column (2). The regressors in columns (3) and (4) are the logarithm of the market access in period 1 (P1 = 1999 - 2001), period 2 (P2 = 2002 - 2004), period 3 (P3 = 2005 - 2007) and period 4 (P4 = 2008 - 2009) calculated with the RV method in column (3) and the HM method in column (4).

Table 5 reports the results for the baseline estimation using both above described methods. Market access is in all cases economically and statistically highly significant and the estimated coefficients are well in line with the results from the directly related research, Breinlich (2006) and Head & Mayer (2006) in particular. Making use of the panel character of the data, columns (1) and (2) report the average impact of market access on wages, with market access calculated with the *RV method* and *HM method* respectively. Columns (3) and (4) allow the coefficient to vary by period. In both specifications, the impact of market access *decreases.* Figure 3 plots the relationship for the period of 1999 -2001. To a large part the downward trend of the coefficient is due to the upward movement of the distinct group of points in the lower left of

Figure 3: GDP per capita and Market Access in 1999 - 2001




*Note:* Average GDP per capita in 1999 - 2001 (2010 EURO) and the regions' market access (LU00 = 100) in 1999 - 2001.

the graph - almost exclusively regions located in the new member states entering the European Union in 2004.[13] The estimated values range between $0.21 - 0.3$ for the *RV method* and $0.28 - 0.38$ for the *HM method*, implying an average increase in GDP per capita between $21\% - 30\%$ and $28\% - 38\%$ respectively for a doubling of the value of market access, depending on the time period and method used. Market access explains between $27\% - 30\%$ of the variation of income. The implied values for $\sigma$, the elasticity of substitution, are plausible as well, when assuming a standard $\alpha = \frac{2}{3}$: they range in between 7.04 and 3.85.

The determination of wages by market access of a region may appear well established through the results, however the attention should be turned to two important issues: First, the use of GDP data, or the estimation of it as through the trade equation, creates an immediate endogeneity problem for the wage equation, where GDP per capita, as a proxy for wages, is regressed on market access. Second, as Redding & Venables (2004) and Breinlich (2006) point out, the returns on the mobile input should equalize across countries and regions by assumption, and would therefore be captured in the constant not the error term, which would create another source of endogeneity.

[13] See table 11 for the corresponding baseline estimation for only EU15 regions and figure 6 for the plot in the period of 2008 - 2009 in appendix C.

Table 6: Instrumental Variable Estimation (1999 - 2001)

| Regressor | ln(MA RV) (1) | ln(GDPpc) (2) | ln(MA HM) (3) | ln(GDPpc) (4) |
|---|---|---|---|---|
| Distance | -1.956e-03*** | | -0.0013882*** | |
| | (-25.05) | | (-16.47) | |
| ln(MA RV) | | 0.24992*** | | |
| | | (8.4192) | | |
| ln(MA HM) | | | | 0.35221*** |
| | | | | (8.4285) |
| Adjusted $R^2$ | 0.7172 | 0.2868 | 0.5224 | 0.3347 |
| Observations | 248 | 248 | 248 | 248 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$. t-statistic in round brackets based on heteroskedasticity-robust standard errors. Due to the nature of the regressor in columns (2) and (4) being itself generated from a prior regression, t-statistics should also be reported based on bootstrapped standard errors, which however could not be done. ln(MA XY) is the logarithm of the market access in the period 1999 - 2001 calculated with the respective method.

### 4.2 Robustness Checks

It is apparent that by construction of the market access variable one should be concerned about endogeneity. The domestic component, even when not explicitly calculated, as in the *RV method*, shows the local expenditure on the considered good. This expenditure is dependent on the income, which itself is of course dependent on the wage, the dependent variable of the estimation, creating a potential reverse causality. As Head & Mayer (2004) note, in the extreme case where transport costs are infinitely high, only local expenditure enters the market access. Redding & Venables (2004) address this concern with excluding the domestic component from the estimation and regressing wages exclusively on what they call *foreign market access*. This in turn however eliminates the arguably most important source of high wages, in particular where market access is high: the demand from the own region. Exemplifying this by taking the Belgium capital region Brussels (NUTS2 code BE10), market access calculated with the *HM method* for the period of 1999 - 2001 is reduced by 53% when excluding the domestic region, while at the same time, the neighboring regions Flemish Brabant (NUTS code BE24) and Walloon Brabant (NUTS code BE31) lose only 5% and 3% respectively, as they draw most of their total market access not from the own region, but from the capital region of Brussels, which is still included in their exclusively non-domestic market access.[14]

Another approach to solve the simultaneity problem is to find a good instrument

[14]This also becomes apparent in table 4, as described above, showing the shares in market access for domestic region, country, Europe and the rest of the world.

Table 7: Lagged Market Access and Fixed Effects

| Regressor | ln(GDPpc) (1) | ln(GDPpc) (2) | ln(GDPpc) (3) | ln(GDPpc) (4) | ln(GDPpc) (5) | ln(GDPpc) (6) |
|---|---|---|---|---|---|---|
| ln(MA RV) | 0.2372*** | | 0.5309*** | | 0.2431*** | |
| | (8.969) | | (5.138) | | (9.784) | |
| ln(MA HM) | | 0.3106*** | | 0.8714*** | | 0.2333*** |
| | | (15.799) | | (27.963) | | (10.335) |
| Time FE | no | no | yes | yes | yes | yes |
| Region FE | no | no | yes | yes | no | no |
| Country FE | no | no | no | no | yes | yes |
| Adjusted $R^2$ | 0.2449 | 0.2616 | 0.9754 | 0.9864 | 0.8371 | 0.8423 |
| Observations | 744 | 744 | 992 | 992 | 992 | 992 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$. t-statistic in round brackets based on heteroskedasticity-robust standard errors. Due to the nature of the regressor being itself generated from a prior regression, t-statistics should also be reported based on bootstrapped standard errors, which however could not be done. ln(MA XY) is the logarithm of the market access calculated with the respective method. The dependent variable is the logarithm of GDP per capita, average per period lagged by one period in columns (1) and (2), and average for the period for columns (3) to (6). Estimations for column (3) and (4) include time and region fixed effects, for column (5) and (6) include time and country fixed effects. Coefficients for all time, region and country fixed effects are not reported.

that is not influenced by wages, but is highly correlated with the calculated market access. As figure 1 shows, the distance to Luxembourg appears to qualify well for this task as a purely geographical variable.[15] Table 6 shows the results for the IV estimation for the period of 1999 - 2001: in the first stage, distance, as the instrumental variable already explains 71% and 52% respectively of variation in regional market access (column 1 for *RV method* and column (2) for *HM method*). In the second stage, the coefficients in columns (2) and (4) are slightly smaller, but they remain highly economically significant and hence very similar in magnitude and are just as significant statistically as in the baseline estimation.

However also the non-domestic component raises concerns: as Breinlich (2006) notes, shocks to wages may be spatially correlated, such as a nationwide strike. To control for such effects, I estimate the same baseline regression with a one-period lag (Table 7, columns (1) and (2)). Country-specific effects that are time-persistent, e.g. institutional setting, can be controlled for by including a country fixed effect (Table 7, columns (5) and (6)). Other unobserved variables may be region-specific and time-

[15] Breinlich (2006) includes a second instrument, country size, to account for a large national market in the light of border effects. In my analysis such measure is insignificant for EU25, but vaguely significant for EU15. Regional size is highly significant with a negative sign, however the endogeneity here is clear: administrative regions are *man-made* smaller in metropolitan areas because of higher population in a smaller area, where income, and hence market access is higher.

persistent (Table 7, columns (5) and (6)). None of these fixed effects or lagged variables reduce the significance of market access for the determination of wages.[16] All controls for robustness underline the importance of market access in the determination of the remuneration for the immobile factor.

## 5 The Role of EU Funds

The European Union Regional Policy, as briefly characterized in section 1, aims to promote cohesion among European regions in economic, social and territorial terms. The means to achieve these forms of cohesion are specified in the objectives, and I here focus on *Objective 1* and the *Cohesion Fund*, and *Objective 2*. The theoretical and econometric framework setup in the previous sections allow to analyze, whether these two objectives have achieved a measurable effect, as they are directly concerned with economic prosperity.

To evaluate the success of this aim in the proposed theoretical framework in section 2, I now analyze the *change in wages*, as postulated in equation (17).

### 5.1 Change in Wages

The estimation of the *change* in wages, is expectedly similar to the estimation of the wage equation (25) in section 4. Recalling equation (17) from section 2:

$$\Delta \ln w_{r,t} = \frac{1}{\alpha\sigma}(\Delta \ln c_t - \Delta \ln a_t) - \frac{\beta}{\alpha}\Delta \ln x_{r,t} + \frac{1}{\alpha\sigma}\Delta \ln MA_{r,t}$$

This is estimated analogous equation to (25) as

$$\ln w_{r,t} - \ln w_{r,t-1} = \theta_t + \frac{1}{\alpha\sigma}(\ln MA_{r,t} - \ln MA_{r,t-1}) + \epsilon_{r,t} \tag{26}$$

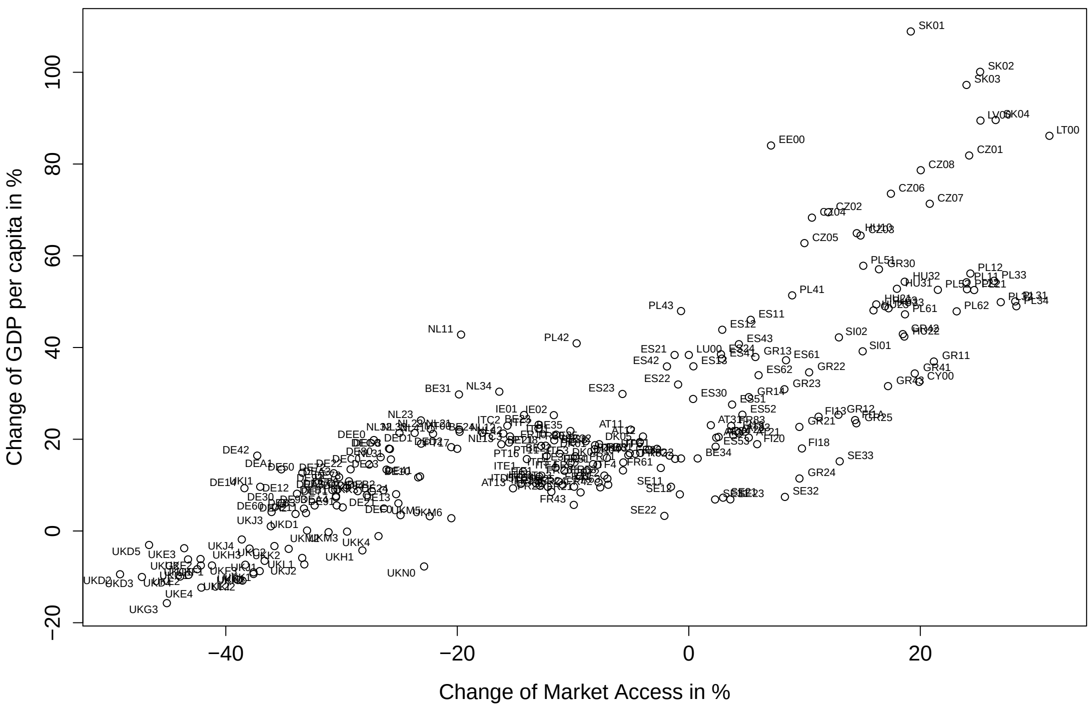
Again $\theta_t$ captures the time-variant $a_t$ and $a_{t-1}$ as well as a change in the elasticity of substitution $c_t$. The return on the mobile factor, $x_{r,t}$ and $x_{r,t-1}$ is captured in the error term $\epsilon_{r,t}$.[17]

In section 5.2 I first estimate equation (26) and control for initial GDP and the change of human capital. Afterwards in section 5.3 I inspect with the estimation of equation (27) whether and how funds from the EU regional policy facilities influenced the change in wages and interacted through the assumed channels.

[16] While the value for the coefficient shoots up when introducing a regional fixed effect (Columns (3) and (4)), in both cases the intercept, not displayed in the table, also drops significantly. The coefficients have no economic interpretation, only their significance is important.

[17] The same possibility of it being captured in the constant as in the equation (25) remains. See section 4.2 for robustness tests in this respect.

Figure 4: Change of Market Access (HM method) and Change in GDP 1999 - 2009




*Note:* Distance to Luxembourg in kilometers and market access relative to Luxembourg (LU00 = 100). Market access calculated with HM method for period from 1999 - 2001.

Data for the EU Regional Funds comes from (Gallo *et al.* , 2011; Dall'erba & Le Gallo, 2008, 2007; Dall'erba, 2005) and EU DG Regio. I follow Lopez-Rodriguez *et al.* (2007) in proxying human capital with by the share of the population aged 25 - 59 (aged 25-64 for 2008) with an educational attainment of at least upper secondary education.[18]

## 5.2 Baseline Estimation of Change in Wages

Columns (1), (2) and (3) of table 8 report the results for the baseline estimation of the change in wages. Column (1) shows that on average for all 248 regions a doubling of the market access over the entire time period from 1999 - 2009 resulted on average in an increase in GDP per capita of 90%. Columns (2) and (3) allow the coefficient to vary between old and new member states, and Objective 1 and 2 regions respectively. The result is revealing: while in the old member states a doubling of the market access of a region resulted in an increase of GDP per capita of only 63%, for new member states it yielded a massive 147%.

[18]Unfortunately data for Denmark, Sweden, Slovenia and the UK Highlands is uncomplete due to a change in NUTS2 regions.

Table 8: Baseline estimation of Change in Wages (HM method, 1999 - 2009), Controls for Initial GDP per capita and Change in Human Capital

| Regressor | Δln(GDPpc) (1) | Δln(GDPpc) (2) | Δln(GDPpc) (3) | Δln(GDPpc) (4) | Δln(GDPpc) (5) | Δln(GDPpc) (6) | Δln(GDPpc) (7) | Δln(GDPpc) (8) | Δln(GDPpc) (9) |
|---|---|---|---|---|---|---|---|---|---|
| Δln(MA) | 0.9073*** | | | 0.7291*** | | | 0.801*** | | |
| | (18.148) | | | (13.3132) | | | (13.3429) | | |
| Δln(MA) EU15 | | 0.6280*** | | | 0.5613*** | | | 0.6402*** | |
| | | (17.6838) | | | (13.1776) | | | (14.8634) | |
| Δln(MA) EU+10 | | 1.46556*** | | | 1.289*** | | | 0.3461 | |
| | | (4.9644) | | | (4.9688) | | | (0.538) | |
| Δln(MA) OBJ1 | | | 0.9781*** | | | 0.9061*** | | | 0.9359*** |
| | | | (13.1154) | | | (10.5078) | | | (10.2015) |
| Δln(MA) OBJ2 | | | 0.7004*** | | | 0.6667*** | | | 0.5423*** |
| | | | (7.2472) | | | (7.7660) | | | (10.6290) |
| initial GDPpc | | | | -5.7e-06*** | -3.6e-06*** | -2.89e-06 | -1.9e-06 | 1.02e-06 | 5.31e-07 |
| | | | | (-3.7895) | (-3.692) | (-1.4957) | (-1.4248) | (1.113) | (0.5195) |
| ΔHC | | | | | | | -0.04061 | | |
| | | | | | | | (-0.1405) | | |
| ΔHC EU15 | | | | | | | | 0.7145** | |
| | | | | | | | | (3.0816) | |
| ΔHC EU+10 | | | | | | | | 0.914 | |
| | | | | | | | | (0.617) | |
| ΔHC OBJ1 | | | | | | | | | -0.5676 |
| | | | | | | | | | (-1.0546) |
| ΔHC OBJ2 | | | | | | | | | 0.8769** |
| | | | | | | | | | (3.2211) |
| EU15 dummy | no | yes | no | no | yes | no | no | yes | no |
| Obj. 1 dummy | no | no | yes | no | no | yes | no | no | yes |
| Adjusted $R^2$ | 0.6455 | 0.7197 | 0.7101 | 0.687 | 0.7335 | 0.7151 | 0.7072 | 0.8168 | 0.7426 |
| Observations | 248 | 248 | 231 | 248 | 248 | 231 | 166 | 166 | 166 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$, t-statistic in round brackets. Due to the nature of the regressor being itself generated from a prior regression, t-statistics should be reported based on bootstrapped standard errors, which however could not be done. The independent variable is the first difference of the logarithm of GDP per capita, as described in the text. The regressors are the difference of the logarithm of Market Access calculated with the HM method, initial GDP per capita, the difference of Human Capital. All differences between the period of 1999 - 2001 and 2008 - 2009. EU15 signals the group of old member states, EU+10 the group of new member states after the enlargement of the EU in 2004. OBJ1 and OB2 group by recipient region of Objective 1 and 2 funds from Regional Policy.

A similar picture is painted when differentiating between Objective 1 and 2 regions:[19] an improvement of market access by 100% for the former yielded a 98% increase in GDP per capita, while it was only 70% for an Objective 2 region. The results hint at the higher impact of market access improvement for lesser developed regions. Figure 4 displays this relationship graphically.

To further establish this relationship, columns (4) to (9) of table 8 control for two standard controls in growth regressions, the change in human capital over the same time period and the initial GDP per capita, as classical growth theory predicts higher growth rates for initially poorer regions.[20] Only controlling for initial GDP expectedly lowers the coefficients in all three specifications, columns (4), (5) and (6), but all remain highly significant. Additionally controlling for the change in human capital shows interesting results, in particular for the specification splitting between Objective 1 and 2 regions in column (9): while market access retains an economic and statistical significance for both groups, the change in human capital is only statistically significant for Objective 2 regions. A 10% increase in human capital in an Objective 2 region yielded on average an increase of 8.8% in GDP per capita. The increase of market access for comparison by 10% yielded only an increase by 5.4% in an Objective 2 region, while boosting GDP per capita in an Objective 1 region by 9.4%. The explained variation of the change in wages varies between $65\% - 82\%$.

### 5.3 Estimation of the Effect of Regional Policy Funds

In the light of these results it is interesting to see whether financial flows from the respective Objective 1 and 2, and Cohesion funds in the programming period of 2000 - 2006 had an influence on the change of wages over the time period of 1999 to 2009.

To estimate this effect I come back to equation (26). The impact of funds from the regional policy can then be estimated with

$$\ln w_{r,t} - \ln w_{r,t-1} = \theta_t + \frac{1}{\alpha\sigma}(\ln MA_{r,t} - \ln MA_{r,t-1}) + H_r EU_r + \epsilon_{r,t} \tag{27}$$

where $EU_r$ is a vector of the different funding sources and $H_r$ a vector of possible interaction channels.

As previously discussed, Objective 1 and the Cohesion Fund have the aim to allow poorer regions to converge, having a primary focus on the financing of infrastructure

[19]Note that, as explained in section 1 Objective 1 regions are comprised of almost all new member state regions, except for Prague, Bratislava and Cyprus, and a smaller number from peripheral countries. The similarity in the direction of the coefficients in columns (2) and (3) is therefore no surprise.

[20]See Mankiw *et al.* (1992) for the seminal work including human capital in a standard neoclassical Solow-Swan model.

projects, while Objective 2 is concerned with better developed regions that face structural problems, having an operative focus on investing in human capital. Relying on the results from the previous section, market access and human capital are in the following taken as proxies for the channels of the two objectives. For the latter the direct link is clear, while for the former, the connection can primarily be drawn by the connection between infrastructure and transport costs in the theoretical model.

For the funds to fulfill their aim, funds from Objective 2 should impact wages through an increase in human capital and for Objective 1 through an increase in market access. However, as Breinlich (2006) and Lopez-Rodriguez *et al.* (2007) show specifically for the case of European regions, human capital itself appears to be one of the major channels of market access to impact wages, making the effects of financial flows from Objective 2 funds difficult to disentangle. This is also visible in table 8 column (9), where a change in market access remains economically and statistically significant for the determination of the change in wages in Objective 2 regions. Likewise it would be plausible that an improvement of market access in Objective 1 regions occurs through the improvement of human capital. In light of the results from section 5.2 this seems not to be the case however, as Objective 1 regions do not appear to be positively influenced by a change in human capital, where the coefficient is statistically insignificant.

The results in table 9 show clearly that a direct connection between the financial flows and a measurable improvement in market access in receiving regions is not present. Including and interacting the respective flows from Objective 1 and 2 facilities with the proposed channels yield (almost) no significant coefficients, neither statistically nor economically. Should there have been a link between market access improvement and financial flows through Objective 1 facilities, the interaction term between both should have exhibited a positive sign and statistical significance. In the case for funds from Objective 2, a positive and significant coefficient would have been expected in *both* interaction terms with market access and human capital, as explained above. This is not the case.

Column (1) reports the results for the estimation with all EU25 countries. To make sure the insignificance of the coefficients was not simply due to omitted variables possibly resulting from the enlargement in 2004, I also estimate the regression with a subsample comprising only the regions of EU15. The result of the lack of a robust connection between changes in market access and human capital and the flows of money from EU Regional Policy facilities clearly does not vanish.

The results are overall *not* surprising in the sense that the funds seem to matter very little. As briefly portrayed in section 1, the economic literature has struggled to establish a conclusive and unconditional favorable link in general.

Table 9: Impact of funds from Objective 1 and Objective 2

| Regressor | Δ ln(GDPpc) (1) | Δ ln(GDPpc) (2) |
|---|---|---|
| Initial GDP per capita | -2.0621e-07 | 2.299e-06* |
| | (-0.1843) | (2.320) |
| Δ MA | 0.30915 | 0.5107* |
| | (0.8852) | (2.174) |
| Δ HC | 1.2213 | 1.128 |
| | (0.9358) | (1.237) |
| ln Obj. 1 | 9.1927e-03 | 2.809e-03 |
| | (1.4734) | (0.669) |
| ln Obj. 2 | 1.8247e-03 | 2.123e-05 |
| | (0.2740) | (0.005) |
| Δ MA x ln Obj. 1 | 2.8943e-02 | 6.042e-03 |
| | (1.6644) | (0.519) |
| Δ HC x ln Obj. 1 | -7.2982e-02 | 3.565e-03 |
| | (-1.0981) | (0.079) |
| Δ MA x ln Obj. 2 | 1.3032e-02 | 1.755e-03 |
| | (0.7075) | (0.145) |
| Δ HC x ln Obj. 2 | -3.2415e-02 | -2.948e-02 |
| | (-0.4600) | -0.610 |
| EU15 only | no | yes |
| Adjusted $R^2$ | 0.733 | 0.716 |
| Observations | 166 | 151 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$. t-statistic in round brackets based on heteroskedasticity-robust standard errors. Due to the nature of the regressor being itself generated from a prior regression, t-statistics should also be reported based on bootstrapped standard errors, which however could not be done. The independent variable is the difference of the logarithms of GDP per capita in the periods of 1999 - 2001 and 2008 - 2009. The regressors are inital GDP per capita in the period of 1999 - 2001, changes in market access and human capital of the time period.Δ MA is the change the market access, calculated with the HM method, and Δ HC the change in human capital. ln Obj. 1 and ln Obj. 2 are the funds allocated to the respective region in per capita terms in the programming period of 2000 - 2006. The lower four rows of regressors display the coefficients for the interaction terms between the concerned variables.

# 6 Conclusion

In this thesis I set up a standard New Economic Geography model in a similar fashion as Redding & Venables (2004) and estimate it with European data. It is closely related to Head & Mayer (2006) and Breinlich (2006) in applying the framework to explain regional differences of income in the European Union. My analysis contributes to this literature by extending the number of inspected regions to 248, encompassing all continental regions of all 25 European Union member states of 2006.

The theoretical framework culminates in two structural equations that make the empirical analysis possible: the *trade equation* and the *wage equation*. I estimate the trade equation with two distinct but closely related approaches that have slightly different requirements of data. One follows closely the original methodology of Redding & Venables (2004) and can be estimated *without* internal trade data. The other approach follows the methodology of Head & Mayer (2004, 2006), where internal trade data is required and has the advantage to capture the home bias in trade. I contrast the results of both methods and highlight the differences rooting in the construction and underlying premisses of the two approaches.

Using the results of the trade equation, I calculate the market access for all regions and estimate the *wage equation*, relating the wage level of a region to its proximity to demand sources. The results are well in line with the findings of other authors and allow a detailed analysis of the Economic Geography of Europe. A clear core-periphery gradient is found. Estimating the first difference of the *wage equation*, also reveals forces that hint at a catching-up of backward regions. A differentiated analysis between old and new member states after the enlargement of the European Union in 2004 highlights this effect. Similarly, the estimation reveals the importance of changes in human capital, in particular in regions that benefit from *Objective 2* of EU Regional Policy. Incomes in *Objective 1* regions are found to be primarily affected by market access, which hints at the importance of basic infrastructure as a prerequisite for economic prosperity.

The wide geographical coverage allows an analysis of the impact of the EU Regional Policy on these previously found determinants of change. However, as previous authors, I cannot find an economically or statistically significant impact of financial facilities eminating from the Objective 1, Cohesion Fund or Objective 2 policies. This result is not a large surprise, as the magnitude of the funds in per capita terms is even in regions with the highest inflows marginal with respect to the time period of 7 years and own economic output.

## References


Anderson, James E., & van Wincoop, Eric. 2004. Trade Costs. *Journal of Economic Literature*, **42**(3), 691–751.

Anderson, J.E., & van Wincoop, E. 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, **93**(1), 170–192.

Becker, Sascha O., Egger, Peter H., & von Ehrlich, Maximilian. 2010. Going NUTS: The effect of EU Structural Funds on regional performance. *Journal of Public Economics*, **94**(9-10), 578–590.

Brakman, Steven, Garretsen, Harry, & Schramm, Marc. 2004. The Spatial Distribution of Wages: Estimating the Helpman-Hanson Model for Germany. *Journal of Regional Science*, **44**(3), 437–466.

Breinlich, Holger. 2006. The spatial income structure in the European Union—what role for Economic Geography? *Journal of Economic Geography*, **6**(5), 593–617.

Chen, Natalie. 2004. Intra-national versus international trade in the European Union: why do national borders matter? *Journal of International Economics*, **63**(1), 93–118.

Combes, Pierre-Philippe, & Overman, Henry G. 2004. The spatial distribution of economic activities in the European Union. *Chap. 64, pages 2845–2909 of:* Henderson, J. V., & Thisse, J. F. (eds), *Handbook of Regional and Urban Economics*. Handbook of Regional and Urban Economics, vol. 4. Elsevier.

Combes, Pierre-Philippe, Lafourcade, M., & Mayer, Thierry. 2005. The trade creating effects of business and social networks: evidence from France. *Journal of International Economics*, **66**, 1–29.

Combes, Pierre-Philippe, Mayer, Thierry, & Thisse, Jaqcues François. 2008. *Economic Geography: The integration of regions and nations*. Princeton University Press.

Crozet, Matthieu. 2004. Do migrants follow market potentials? An estimation of a new economic geography model. *Journal of Economic Geography*, **4**(4), 439–458.

Crozet, Matthieu, & Lafourcade, Miren. 2009. *La nouvelle économie géographique*. la Découverte.

Dall'erba, Sandy. 2005. Distribution of Regional Income and Regional Funds in Europe 1989-1999: an Exploratory Spatial Data Analysis. *The Annals of Regional Science*, **39**(1), 121–148.

Dall'erba, Sandy. 2008. European Cohesion Policy - Willem Molle. *Journal of Economic Geography*, **8**(4), 587–588.

Dall'erba, Sandy, & Le Gallo, Julie. 2007. Cohesion policy, the Convergence Process and Employment in the European Union. *Czech Journal of Economics and Finance*, **57**(7), 324–340.

Dall'erba, Sandy, & Le Gallo, Julie. 2008. Regional convergence and the impact of European structural funds over 1989–1999: A spatial econometric analysis*. *Papers in Regional Science*, **87**(2), 219–244.

Davis, Donald R., & Weinstein, David E. 2002. Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review*, **92**(5), 1269–1289.

Disdier, Anne-Célia, & Head, Keith. 2008. The Puzzling Persistence of the Distance Effect on Bilateral Trade. *The Review of Economics and Statistics*, **90**(1), 37–48.

European Commission. 2008a. *EU Cohesion Policy 1988-2008: Investing in Europe's future.*

European Commission. 2008b. *Regions of the European Union - A Statistical Portrait - 2009 Edition.* Office for Official Publications of the European Communities.

European Commission. 2008c. *Working for the Regions - EU Regional Policy 2007-2013.* European Commission, Directorate-General for Regional Policy.

European Communities. 2004. *Working for the Regions.* Office for Official Publications of the European Communities, 2004.

Fujita, Masahisa, Krugman, Paul, & Venables, Anthony J. 1999. *The Spatial Economy: Cities, Regions, and International Trade.* MIT Press Books, vol. 1, no. 0262561476. The MIT Press.

Gallo, Julie Le, Dall'Erba, Sandy, & Guillain, Rachel. 2011. The Local versus Global Dilemma of the Effects of Structural Funds. *Growth and Change*, **42**(4), 466–490.

Gaulier, Guillaume, & Zignago, Soledad. 2010. *BACI: International Trade Database at the Product-level, The 1994-2007 Version.* Tech. rept. CEPII.

Hanson, Gordon H. 1998 (Feb.). *Market Potential, Increasing Returns, and Geographic Concentration.* NBER Working Papers 6429. National Bureau of Economic Research, Inc.

Hanson, Gordon H. 2005. Market potential, increasing returns and geographic concentration. *Journal of International Economics*, **67**(1), 1 – 24.

Head, Keith, & Mayer, Thierry. 2002 (Jan.). *Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade.* Working Papers 2002-01. CEPII research center.

Head, Keith, & Mayer, Thierry. 2004. Market Potential and the Location of Japanese Investment in the European Union. *The Review of Economics and Statistics*, **86**(4), 959–972.

Head, Keith, & Mayer, Thierry. 2006. Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics*, **36**(5), 573 – 594.

Head, Keith, & Mayer, Thierry. 2011. Gravity, market potential and economic development. *Journal of Economic Geography*, **11**(2), 281–294.

Head, Keith, Mayer, Thierry, & Ries, John. 2010. The erosion of colonial trade linkages after independence. *Journal of International Economics*, **81**(1), 1–14.

Hering, Laura, & Poncet, Sandra. 2010. Income Per Capita Inequality in China: The Role of Economic Geography and Spatial Interactions. *World Economy*, **33**(5), 655–679.

Krugman, Paul. 1991. Increasing Returns and Economic Geography. *Journal of Political Economy*, **99**(3), 483–99.

Lopez-Rodriguez, Jesus, Faina, J. Andres, & Lopez-Rodriguez, Jose. 2007. Human Capital Accumulation and Geography: Empirical Evidence from the European Union. *Regional Studies*, **41**(2), 217–234.

Mankiw, N. Gregory, Romer, David, & Weil, David N. 1992. A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, **107**(2), 407 – 437.

Mayer, Thierry, & Zignago, Soledad. 2011. *Notes on CEPII's distances measures: The GeoDist database*. Tech. rept. CEPII.

Mundell, R.A. 1961. A theory of optimum currency areas. *The American Economic Review*, **51**(4), 657–665.

Paillacar, Rodrigo. 2009 (July). *An empirical study of the world economic geography of manufacturing industries (1980 - 2003)*. unpublished.

Redding, Stephen, & Schott, Peter K. 2003. Distance, skill deepening and development: will peripheral countries ever get rich? *Journal of Development Economics*, **72**(2), 515–541.

Redding, Stephen, & Venables, Anthony J. 2004. Economic geography and international inequality. *Journal of International Economics*, **62**(1), 53–82.

Redding, Stephen J., & Sturm, Daniel M. 2008. The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review*, **98**(5), 1766–97.

Rodriguez-Pose, Andres, & Fratesi, Ugo. 2004. Between Development and Social Policies: The Impact of European Structural Funds in Objective 1 Regions. *Regional Studies*, **38**(1), 97–113.

Treaty of Rome. 1957. *Treaty of Rome: Treaty Establishing the European Economic Community.*

Wei, Shang-Jin. 1996 (Apr.). *Intra-National versus International Trade: How Stubborn are Nations in Global Integration?* NBER Working Papers 5531. National Bureau of Economic Research, Inc.

Wolf, Holger C. 2000. Intranational Home Bias In Trade. *The Review of Economics and Statistics*, **82**(4), 555–563.

# A Appendix: Regional Geographic Data

Geographic data for European regions is assembled from a number of sources and then merged and constructed analogous to Mayer & Zignago (2011).

Regions are NUTS2 regions in the categorization of 2006. Area and population size data is taken from ESPON ReRisk Area Population Data (years 2005 and 2003 for UK regions). The capital of a region is assumed to be its centroid. In region that have no administrative capital, the largest city by population size is assumed to be the centroid (This is the case only for the UK, the Czech Republic and Slovakia). The centroid of a region that fully encompasses a smaller region has the same centroid as this enclave (This is the case only for the metropolitan areas of London, UKI1 and UKI2, and the greater Prague area, CZ01 and CZ02). Latitude and longitude are coded using the Google Data API, distances are calculated as simple great circle distances, following Paillacar (2009) despite criticism from Head & Mayer (2002) who propose a population weighted distance calculated as $d_{ij} = \sum_{m\epsilon i} \sum_{n\epsilon j} \frac{pop_m}{pop_i} \frac{pop_n}{pop_j} dist_{mn}$. Internal distance of a region are calculated as i $d_{ii} = \frac{2}{3}\sqrt{area/\pi}$ following (Mayer & Zignago, 2011; Head & Mayer, 2002). Regional official regional languages come from DG Regio and the CIA Factbook.

NUTS2 regions considered: AT11 (Burgenland (A)), AT12 (Niederösterreich), AT13 (Wien), AT21 (Kärnten), AT22 (Steiermark), AT31 (Oberösterreich), AT32 (Salzburg), AT33 (Tirol), AT34 (Vorarlberg), BE10 (Brussels Hoofdstedelijk Gewest), BE21 (Prov. Antwerpen), BE22 (Prov. Limburg (B)), BE23 (Prov. Oost-Vlaanderen), BE24 (Prov. Vlaams Brabant), BE25 (Prov. West-Vlaanderen), BE31 (Prov. Brabant Wallon), BE32 (Prov. Hainaut), BE33 (Prov. Liège), BE34 (Prov. Luxembourg (B)), BE35 (Prov. Namur), CY00 (Cyprus), CZ01 (Praha), CZ02 (Strední Cechy), CZ03 (Jihozápad), CZ04 (Severozápad), CZ05 (Severov_chod), CZ06 (Jihov_chod), CZ07 (Strední Morava), CZ08 (Moravskoslezsko), DE11 (Stuttgart), DE12 (Karlsruhe), DE13 (Freiburg), DE14 (Tübingen), DE21 (Oberbayern), DE22 (Niederbayern), DE23 (Oberpfalz), DE24 (Oberfranken), DE25 (Mittelfranken), DE26 (Unterfranken), DE27 (Schwaben), DE30 (Berlin), DE41 (Brandenburg - Nordost), DE42 (Brandenburg - Südwest), DE50 (Bremen), DE60 (Hamburg), DE71 (Darmstadt), DE72 (Gießen), DE73 (Kassel), DE80 (Mecklenburg-Vorpommern), DE91 (Braunschweig), DE92 (Hannover), DE93 (Lüneburg), DE94 (Weser-Ems), DEA1 (Düsseldorf), DEA2 (Köln), DEA3 (Münster), DEA4 (Detmold), DEA5 (Arnsberg), DEB1 (Koblenz), DEB2 (Trier), DEB3 (Rheinhessen-Pfalz), DEC0 (Saarland), DED1 (Chemnitz), DED2 (Dresden), DED3 (Leipzig), DEE0 (Sachsen-Anhalt), DEF0 (Schleswig-Holstein), DEG0 (Thüringen), DK01 (Hovedstaden), DK02 (Sjælland), DK03 (Syddanmark), DK04 (Midtjylland), DK05 (Nordjylland), EE00 (Estonia), ES11 (Galicia), ES12 (Principado de Asturias), ES13 (Cantabria), ES21 (Pais Vasco), ES22 (Comunidad Foral de Navarra), ES23 (La Rioja), ES24 (Aragón), ES30 (Comunidad de Madrid), ES41

(Castilla y León), ES42 (Castilla-la Mancha), ES43 (Extremadura), ES51 (Cataluña), ES52 (Comunidad Valenciana), ES53 (Illes Balears), ES61 (Andalucia), ES62 (Región de Murcia), FI13 (Itä-Suomi), FI18 (Etelä-Suomi), FI19 (Länsi-Suomi), FI1A (Pohjois-Suomi), FI20 (Åland), FR10 (Île de France), FR21 (Champagne-Ardenne), FR22 (Picardie), FR23 (Haute-Normandie), FR24 (Centre), FR25 (Basse-Normandie), FR26 (Bourgogne), FR30 (Nord - Pas-de-Calais), FR41 (Lorraine), FR42 (Alsace), FR43 (Franche-Comté), FR51 (Pays de la Loire), FR52 (Bretagne), FR53 (Poitou-Charentes), FR61 (Aquitaine), FR62 (Midi-Pyrénées), FR63 (Limousin), FR71 (Rhône-Alpes), FR72 (Auvergne), FR81 (Languedoc-Roussillon), FR82 (Provence-Alpes-Côte d'Azur), FR83 (Corse), GR11 (Anatoliki Makedonia, Thraki), GR12 (Kentriki Makedonia), GR13 (Dytiki Makedonia), GR14 (Thessalia), GR21 (Ipeiros), GR22 (Ionia Nisia), GR23 (Dytiki Ellada), GR24 (Sterea Ellada), GR25 (Peloponnisos), GR30 (Attiki), GR41 (Voreio Aigaio), GR42 (Notio Aigaio), GR43 (Kriti), HU10 (Közép-Magyarország), HU21 (Közép-Dunántúl), HU22 (Nyugat-Dunántúl), HU23 (Dél-Dunántúl), HU31 (Észak-Magyarország), HU32 (Észak-Alföld), HU33 (Dél-Alföld), IE01 (Border, Midlands and Western), IE02 (Southern and Eastern), ITC1 (Piemonte), ITC2 (Valle d'Aosta/Vallée d'Aoste), ITC3 (Liguria), ITC4 (Lombardia), ITD1 (Provincia Autonoma Bolzano-Bozen), ITD2 (Provincia Autonoma Trento), ITD3 (Veneto), ITD4 (Friuli-Venezia Giulia), ITD5 (Emilia-Romagna), ITE1 (Toscana), ITE2 (Umbria), ITE3 (Marche), ITE4 (Lazio), ITF1 (Abruzzo), ITF2 (Molise), ITF3 (Campania), ITF4 (Puglia), ITF5 (Basilicata), ITF6 (Calabria), ITG1 (Sicilia), ITG2 (Sardegna), LT00 (Lithuania), LU00 (Luxembourg), LV00 (Latvia), MT00 (Malta), NL11 (Groningen), NL12 (Friesland (NL)), NL13 (Drenthe), NL21 (Overijssel), NL22 (Gelderland), NL23 (Flevoland), NL31 (Utrecht), NL32 (Noord-Holland), NL33 (Zuid-Holland), NL34 (Zeeland), NL41 (Noord-Brabant), NL42 (Limburg (NL)), PL11 (Lódzkie), PL12 (Mazowieckie), PL21 (Malopolskie), PL22 (Slaskie), PL31 (Lubelskie), PL32 (Podkarpackie), PL33 (Swietokrzyskie), PL34 (Podlaskie), PL41 (Wielkopolskie), PL42 (Zachodniopomorskie), PL43 (Lubuskie), PL51 (Dolnoslaskie), PL52 (Opolskie), PL61 (Kujawsko-Pomorskie), PL62 (Warminsko-Mazurskie), PL63 (Pomorskie), PT11 (Norte), PT15 (Algarve), PT16 (Centro (PT)), PT17 (Lisboa), PT18 (Alentejo), SE11 (Stockholm), SE12 (Östra Mellansverige), SE21 (Småland med öarna), SE22 (Sydsverige), SE23 (Västsverige), SE31 (Norra Mellansverige), SE32 (Mellersta Norrland), SE33 (Övre Norrland), SI01 (Vzhodna Slovenija), SI02 (Zahodna Slovenija), SK01 (Bratislavsk_ kraj), SK02 (Západné Slovensko), SK03 (Stredné Slovensko), SK04 (V_chodné Slovensko), UKC1 (Tees Valley and Durham), UKC2 (Northumberland, Tyne and Wear), UKD1 (Cumbria), UKD2 (Cheshire), UKD3 (Greater Manchester), UKD4 (Lancashire), UKD5 (Merseyside), UKE1 (East Yorkshire and Northern Lincolnshire), UKE2 (North Yorkshire), UKE3 (South Yorkshire), UKE4 (West Yorkshire), UKF1 (Derbyshire and Nottinghamshire), UKF2 (Leicestershire, Rutland and Northants), UKF3 (Lincolnshire), UKG1 (Herefordshire, Worcestershire and Warks), UKG2 (Shropshire and Staffordshire), UKG3

(West Midlands), UKH1 (East Anglia), UKH2 (Bedfordshire, Hertfordshire), UKH3 (Essex), UKI1 (Inner London), UKI2 (Outer London), UKJ1 (Berkshire, Bucks and Oxfordshire), UKJ2 (Surrey, East and West Sussex), UKJ3 (Hampshire and Isle of Wight), UKJ4 (Kent), UKK1 (Gloucestershire, Wiltshire and Bristol/Bath area), UKK2 (Dorset and Somerset), UKK3 (Cornwall and Isles of Scilly), UKK4 (Devon), UKL1 (West Wales and The Valleys), UKL2 (East Wales), UKM2 (Eastern Scotland), UKM3 (South Western Scotland), UKM5 (North Eastern Scotland), UKM6 (Highlands and Islands), UKN0 (Northern Ireland).

# B Construction of Variables

GDP per capita is Gross domestic product (GDP) at current market prices at NUTS level 2 taken from EUROSTAT. GDP for the years 1995 - 2007 for Austria, Italy and Hungary is from a 2010 version from EUROSTAT as this data is unavailable currently.

Manufacturing production data is taken from the OECD STAN database (PROD: PROD Production (gross output), current prices, C15T37 Manufacturing)

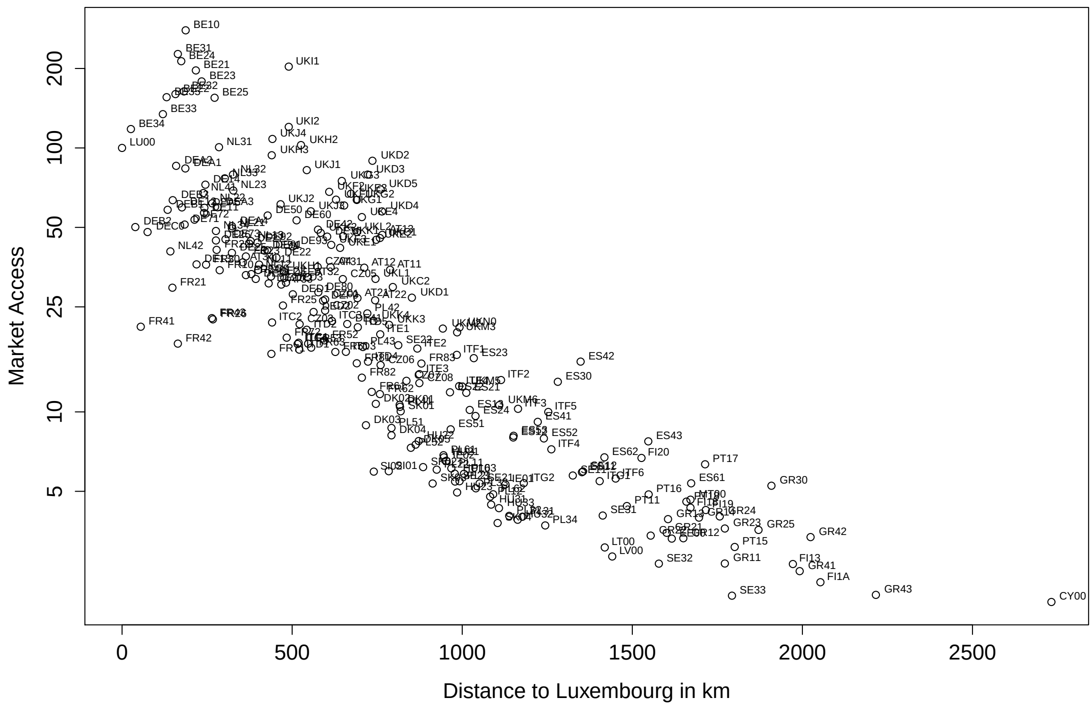
All monetary values are in 2010 Euros, USD are converted to 2010 EUR using ECB annual average data.

Table 10: Summary Statistics for the Data

| Variable | Observations | Mean | Std. Dev. |
|---:|---:|---:|---:|
| Trade flow | 358685 | 306400 | 7492606 |
| Distance | 60450 | 6211.00 | 4465.041 |
| GDP per capita | 248 | 23556 | 10756.16 |
| Objective 1 + Cohesion funds | 225 | 537400000 | 1114301777 |
| Objective 1 + Cohesion funds per capita | 225 | 268.2 | 509.3136 |
| Objective 2 | 225 | 88960000 | 138845167 |
| Objective 2 per capita | 225 | 51.17 | 64.71 |
| Population Share with upper secondary education or higher | 812 | 23.00 | 8.01267 |
| Population | 1736 | 1839.0 | 1508.797 |

# C Further tables and plots

Figure 5: Distance to Luxembourg and Market Access in Period 1999 - 2001



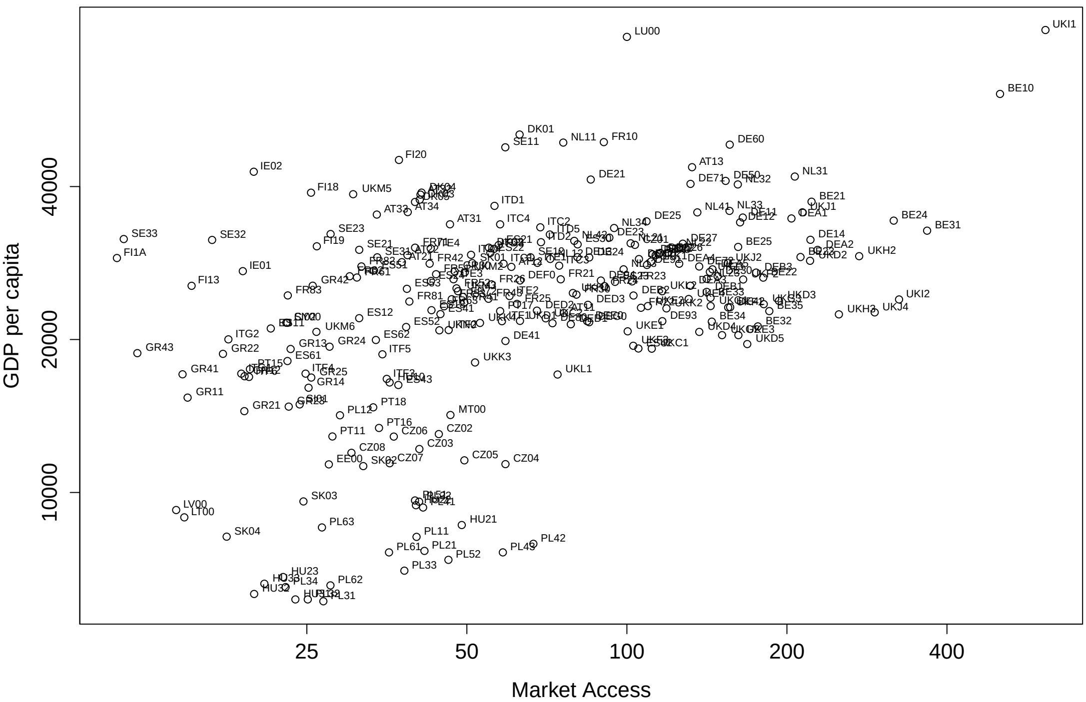

*Note:* Distance to Luxembourg in kilometers and market access relative to Luxembourg (LU00 = 100). Market access calculated with RV method for period from 1999 - 2001.

Table 11: Baseline estimation wage equation 1999 - 2009 for EU15 only

| Regressor | ln(GDP p.c.) | ln(GDP p.c.) | ln(GDP p.c.) | ln(GDP p.c.) |
|---|---|---|---|---|
| ln(MA RV) | 0.148231*** | | | |
| | (13.8169) | | | |
| ln(MA HM) | | 0.17547*** | | |
| | | (13.4664) | | |
| ln(MA) P1 | | | 0.170570*** | 0.19565*** |
| | | | (8.0091) | (8.3155) |
| ln(MA) P2 | | | 0.1489543*** | 0.1762672*** |
| | | | (7.0348) | (6.9433) |
| ln(MA) P3 | | | 0.1401249*** | 0.1750146*** |
| | | | (6.5190) | (6.4872) |
| ln(MA) P4 | | | 0.130851*** | 0.147507*** |
| | | | (6.0367) | (5.0951) |
| EU15 only | yes | yes | yes | yes |
| Period Dummies | yes | yes | yes | yes |
| Adjusted $R^2$ | 0.2989 | 0.2988 | 0.2722 | 0.2694 |
| Observations | 792 | 792 | 792 | 792 |
| Year | 1999 - 2009 | 1999 - 2009 | 1999 - 2009 | 1999 - 2009 |

*Note:* statistical significance * at $p < 0.1$, ** at $p < 0.05$, *** at $p < 0.01$, t-statistic in round brackets. Due to the nature of the regressor being itself generated from a prior regression, t-statistics should be reported based on bootstrapped standard errors, which however could not be done. The independent variable is the logarithm of the mean of GDP per capita in the respective period, as described in the text. The regressors in columns (1) and (2) are the logarithms of the market access calculated with the RV method in column (1) and HM method in column (2). The regressors in columns (3) and (4) are the logarithm of the market access in period 1 (P1 = 1999 - 2001), period 2 (P2 = 2002 - 2004), period 3 (P3 = 2005 - 2007) and period 4 (P4 = 2008 - 2009) calculated with the RV method in column (3) and the HM method in column (4).

Figure 6: GDP per capita and Market Access in 2008 - 2009




*Note:* Average GDP per capita in 2008 - 2009 (2010 EURO) and the regions' market access (LU00 = 100) in 2008 - 2009.